MicroRNAs in HPV-induced cervical cancer

Babion, I.

2020

document version
Publisher's PDF, also known as Version of record

Genome-Wide miRNA Analysis of HPV-Positive Self-Samples Yields Novel Triage Markers for Early Detection of Cervical Cancer

Barbara C. Snoek
Iris Babion*
Wina Verlaat*
Putri W. Novianti
Mark A. van de Wiel
Saskia M. Wilting
Nienke E. van Trommel
Maaike C.G. Bleeker
Leon F.A.G. Massuger
Willem J.G. Melchers
Daoud Sie
Daniëlle A.M. Heideman
Peter J.F. Snijders
Chris J.L.M. Meijer
Renske D.M. Steenbergen

* Both authors contributed equally.

International Journal of Cancer 2019; 144(2):372-379

ABSTRACT

Background

Offering self-sampling for hrHPV testing improves the effectiveness of current cervical screening programs by increasing population coverage. Molecular markers directly applicable on self-samples are needed to stratify hrHPV-positive women at risk of cervical cancer (so-called triage) and to avoid over-referral and overtreatment. Deregulated microRNAs (miRNAs) have been implicated in the development of cervical cancer and represent potential triage markers. However, it is unknown whether deregulated miRNA expression is reflected in self-samples. Our study is the first to establish genome-wide miRNA profiles in hrHPV-positive self-samples to identify miRNAs that can predict the presence of CIN3 and cervical cancer in self-samples.

Methods

Small RNA sequencing (sRNA-Seq) was conducted to determine genome-wide miRNA expression profiles in 74 hrHPV-positive self-samples from women with and without cervical precancer (CIN3). The optimal miRNA marker panel for CIN3 detection was determined by GRridge, a penalized method on logistic regression. Six miRNAs were validated by qPCR in 191 independent hrHPV-positive self-samples.

Results

Classification of sRNA-Seq data yielded a 9-miRNA marker panel with a combined area under the curve (AUC) of 0.89 for CIN3 detection. Validation by qPCR resulted in a combined AUC of 0.78 for CIN3+ detection.

Conclusion

Our study shows that deregulated miRNA expression associated with CIN3 and cervical cancer development can be detected by sRNA-Seq in hrHPV-positive self-samples. Validation by qPCR indicates that miRNA expression analysis offers a promising novel molecular triage strategy for CIN3 and cervical cancer detection applicable to self-samples.

INTRODUCTION

In many countries, the participation of women in population-based cervical screening is suboptimal, which hampers the effectiveness of the screening program[1]. Women who do not participate (30% of invited women in The Netherlands), so-called non-attendees, are at increased risk of developing cervical intraepithelial neoplasia grade 3 or worse (CIN3+)[2]. Offering self-sampling for high-risk human papillomavirus (hrHPV) testing to non-attendees has been shown to increase the efficacy of the screening program[3] and has demonstrated to be equally effective in detecting CIN3+ when compared with hrHPV testing on physician-taken specimens[4,5]. Therefore, self-sampling for hrHPV testing is now offered to non-attendees in The Netherlands. To discriminate between women with a transient (clinically irrelevant) hrHPV infection and women with clinically meaningful cervical disease, additional risk assessment (so-called triage) is necessary. For cervical scrapes, testing of hrHPV-positive women by (repeat) cytology is a generally acknowledged triage strategy[6,7]. Since cytological examination of self-samples is not reliable[8], women testing positive for hrHPV on their self-sample need to have a physician-taken scrape for triage testing by cytology. This prerequisite for cervical cytology for previous non-attendees has shown to be associated with loss to follow-up[9]. Therefore, the identification and subsequent use of non-morphological, molecular triage markers that can be directly applied on self-samples become very important.

Molecular triage methods that have been described so far for application on self-samples include HPV16/18 genotyping[10,11] and DNA methylation marker analysis[9,10,12,13]. MicroRNAs (miRNAs) have emerged as key regulators of gene expression and their involvement in carcinogenesis has now been widely accepted[14]. They are increasingly recognized as promising biomarkers for cancer diagnostics and may provide alternative molecular triage markers. Interestingly, Tian *et al*. reported that miRNA detection in hrHPV-positive cervical scrapes has a superior performance over cytology for the detection of high-grade CIN lesions in a referral population[15], which indicates the potential of miRNA detection for triage testing of hrHPV-positive women in cervical screening programs.

A number of studies have demonstrated altered miRNA expression patterns in cervical cancer cell lines and cervical cancer tissues[16]. However, little is known about genome-wide miRNA expression patterns in CIN lesions[17,18] and no data is presently available on miRNA expression profiles in hrHPV-positive self-

samples. Importantly, previous research has shown that the clinical performance of molecular markers may be dependent on the sample type used[12], given that the cellular composition may differ between tissues, physician-taken cervical scrapes and self-samples and the fact that non-disease-related cells, such as vaginal cells, are largely over-represented in self-samples. Accordingly, the optimum miRNAs listed so far may not be directly extrapolated to self-samples.

Our study set out to directly determine the genome-wide miRNA profiles in self-samples aiming to establish a miRNA signature for CIN3 detection to serve as a novel triage method for hrHPV-positive self-samples. To this end, we performed small RNA next generation sequencing (sRNA-Seq) on hrHPV-positive self-samples from women with CIN3 lesions and women without cervical disease in follow-up (≤ CIN1), obtained from a prospective screening trial among non-attendees[3]. Penalized logistic regression analysis resulted in the identification of a panel of nine miRNAs that together could detect CIN3 with high discriminatory power. The performance of this miRNA signature was subsequently validated by qPCR in an independent sample series of hrHPV-positive self-samples.

MATERIALS AND METHODS

Clinical Cervical Specimens and RNA Extraction

For the discovery set, we used 74 hrHPV-positive cervico-vaginal self-samples that were collected from non-attendees participating in the PROHTECT-1 trial (NTR792)[3] with the first generation Delphi Screener (Delphi Bioscience, The Netherlands) between December 2006 and December 2007. Detailed characteristics of study design and clinical specimens, inclusion criteria and follow-up procedures have been described previously[3]. MiRNA sequencing data from a pilot experiment of 12 self-samples for power calculations revealed a ratio of 1 (hrHPV-positive controls) to 1 (CIN3) for proper marker discovery. Therefore, the discovery set consisted of 36 control women with either histologically confirmed ≤ CIN1 or that displayed hrHPV clearance combined with normal cytology in follow-up (hereafter referred to as controls; median age of 37; range 31-56) and 38 women who were histologically diagnosed with a CIN3 lesion (hereafter referred to as cases; median age of 39; range 31-62). CIN2 lesions were not included given their ambiguous classification; they often represent a misclassified CIN1 or CIN3[19]. Controls and cases were matched according to age and hrHPV type. Total RNA was isolated using Direct-Zol™ RNA Miniprep (Zymo Research, Freiburg, Germany) after the manufacturer's instructions.

For the validation set, we used 155 hrHPV-positive lavage self-samples that were collected from non-attendees participating in the methylation triage arm of the PROHTECT-3 trial (NTR2606)[9] with the second generation Delphi Screener (Delphi Bioscience) between November 2010 and December 2011. Detailed characteristics of study design and clinical specimens, inclusion criteria, and follow-up procedures have been described previously[9]. These 155 hrHPV-positive self-samples consisted of 101 control women with either histologically confirmed ≤ CIN1 or that displayed hrHPV clearance combined with normal cytology in follow-up (hereafter referred to as controls; median age of 42; range 33-63), 49 women who were histologically diagnosed with a CIN3 lesion (median age of 40; range 33-58) and five women who were histologically diagnosed with squamous cell carcinoma (SCC; median age of 47; range 38-58). Furthermore, 36 hrHPV-positive brush-based self-samples (median age of 46; range 27-83) were included that were obtained from women who visited the Antoni van Leeuwenhoek Hospital/Netherlands Cancer Institute or Academic Medical Centre Amsterdam for treatment of SCC between 2015 and 2017 (XMET and SOLUTION study). Total RNA was isolated using Trizol reagent (Thermo Fisher Scientific, Bleiswijk, The Netherlands) according to the manufacturer's instructions. Samples from both PROHTECT trials were tested for hrHPV by GP5+/6+ PCR using the Diassay EIA HPV GP HR kit (Diassay, Voorburg, The Netherlands) as described previously[9,20]. Samples from the XMET and SOLUTION studies were tested for hrHPV using the HPV-Risk Assay (Self-screen BV, Amsterdam, The Netherlands)[21].

Ethical approval was obtained from the National Health Council (PROHTECT trials) and Institutional Review Board of the Antoni van Leeuwenhoek Hospital/Netherlands Cancer Institute (METC15.1468) and VU University Medical Centre (METC2016.213). All participants gave informed consent. All samples were used in an anonymous fashion in accordance with the "Code for Proper Secondary Use of Human Tissues in the Netherlands" as formulated by the Dutch Federation of Medical Scientific Organisations (http://www.fmwv.nl and www.federa.org)[22].

Library Construction and Deep Sequencing

Preparation of cDNA libraries was performed with equal amounts of input RNA per sample (1000ng of total RNA) using the TruSeq Small RNA Sample Preparation Kit according to the manufacturer's instruction (Illumina, San Diego, CA). In brief, cDNA libraries were measured on an Agilent 2100 Bioanalyzer (Agilent Technologies) and up to five samples were pooled in equimolar concentrations for the sequencing run. Sequencing was performed using single read 50 (SR50) cycles on a HiSeq 2500 (Illumina) for 74 samples (36 controls, 38 CIN3).

Data Processing miRNA Sequencing

In order to obtain miRNA expression profiles, all raw small RNA sequencing data was analyzed using the miRDeep2 algorithm[23,24]. In brief, read counts were filtered for low-quality sequences, adapters, and sequences shorter than 18 nucleotides followed by alignment to the human genome (hg19). Input from known miRNAs was obtained from miRBase version 20[25]. To avoid multiple counting of miRNA sequences that can originate from several genomic loci, we only counted the expression of miRNA paralogs once. As part of our quality control, we assessed the distribution of miRNA read counts and excluded those samples that depicted an atypical distribution (Fig. S1). Based on this criterion, 10 samples with insufficient quality were excluded. Another eight samples were already excluded beforehand because of low sequencing depth.

After quality control, miRNA read counts were normalized according to the trimmed mean of M values (TMM) normalization strategy[26]. Subsequently, normalized miRNA read counts were square root-transformed to the quasi Gaussian scale followed by miRNA data standardization. Finally, from 2,577 known miRNAs according to miRBase 20, only those miRNAs that exhibited non-zero read counts in at least three samples were included. This resulted in a total set of 772 miRNAs with an average of 416 miRNAs per sample (controls: 403, CIN3: 434). Raw sequencing reads and quantified read counts are available from the NCBI Gene Expression Omnibus (GEO) through series accession number GSE104758.

Statistical Analysis miRNA Sequencing and Microarray Data

Adaptive group regularized (logistic-) ridge regression (GRridge)[27] was applied to the preprocessed miRNA sequencing data (expression values). Auxiliary information, namely abundance and conservation status[28], was incorporated to build the omics-based prediction model. ***Post hoc*** feature selection was subsequently applied to the GRridge model, which resulted in the identification of nine miRNAs to be further validated by quantitative PCR (qPCR). More details of the GRridge predictive modelling as well as the results (including data preprocessing and predictive modelling) have been described elsewhere[29]. The performance of the GRridge prediction model on the miRNA expression data was assessed by leave-one-out cross-validation (LOOCV), then visualized by the receiver operating characteristics (ROC) curve, and quantified by area under the curve (AUC).

In-house available genome-wide microarray miRNA profiles of cervical tissues[30], including normal cervical tissues, CIN2/3 lesions (hereafter referred to as high-grade CIN), and SCC were utilized to evaluate the expression patterns of the selected miRNAs. Due to limitations in accurate histologic grading of high-grade CIN on frozen tissue specimens, CIN2 and CIN3 were grouped together in this sample series[30]. Since the miRNA microarray did not include probes for miR-184, the expression of miR-184 in cervical tissues could not be evaluated. For the selected miRNAs, we employed the Kruskal-Wallis omnibus test and the *post hoc* non-parametric test to compare miRNA expression between biological groups. Obtained p-values were adjusted for multiple comparisons. Statistical analyses were performed in R open source software by implementing the GRridge[27] package.

Quantitative RT-PCR (qPCR)

Expression of hsa-let-7b-5p, hsa-miR-9-5p, hsa-miR-15b-5p, hsa-miR-20a-5p, hsa-miR-31-5p, hsa-miR-93-5p, hsa-miR-183-5p, hsa-miR-184, and hsa-miR-222-3p was determined using TaqMan microRNA assays (002619, 000583, 000390, 000580, 002279, 001090, 002269, 000485, 002276; Thermo Fisher Scientific). Since we most recently described hsa-miR-423-3p and hsa-miR-30b-5p as suitable normalization strategy for hrHPV-positive self-samples[31], these miRNAs were included as reference genes (002626, 000602; Thermo Fisher Scientific).

The manufacturer's protocol was adapted in order to multiplex reverse transcription of all nine miRNA markers and two reference genes (11 miRNAs in total) and validated in comparison to singleplex reverse transcription reactions. In short, the specific reverse transcription primers were combined in a primer pool and cDNA was synthesized in 16.5µl reactions containing 9µl primer pool, 0.45µl dNTPs (100 mM), 2.25µl 10× RT buffer, 0.3µl RNase inhibitor (20U/µl), and 4.5µl MultiScribe Reverse Transcriptase (TaqMan microRNA Reverse Transcription kit, Thermo Fisher Scientific). We used 50ng total RNA as reverse transcription template.

Quantitative PCR reactions were performed in 10µl, consisting of 5µl TaqMan® Universal Master Mix II no UNG, 0.5µl miRNA specific TaqMan assays (Thermo Fisher Scientific), 3.5µl H_2O, and 1µl cDNA. The ABI7500 Fast Real-Time PCR System (Thermo Fisher Scientific) was used for quantification. Cycle conditions used for cDNA synthesis and PCR were according to the manufacturer's protocols. Samples that exhibited C_t values > 32 for the geometric mean of the

reference genes were considered unsuitable for miRNA qPCR analysis. Based on this criterion, one CIN3 self-sample was excluded. For the remaining self-samples, miRNA expression was normalized to the geometric mean of the reference genes applying the $2^{-\Delta Ct}$ method[32].

Detailed characteristics of physician-taken cervical scrapes, including miRNA expression analysis by qPCR have been described previously[33].

Statistical Analysis miRNA qPCR Data

Since our panel of miRNA markers should be able to detect cervical precancerous lesions with increased risk to progress to cervical cancer as well as cervical cancer with high discriminatory power, we combined CIN3 and cervical cancer samples (CIN3+) for the analysis of the validation set. Logistic regression analysis was performed to assess the prediction ability of the six miRNAs on the normalized qPCR data for detection of CIN3+. First, the C_t ratio of each miRNA was transformed to the square root scale. Next, a univariable logistic regression model was fitted to observe the performance of each miRNA. Multivariable logistic regression with backward elimination was subsequently applied to define the best combination of miRNAs to differentiate self-samples from women with CIN3+ from hrHPV-positive controls. Table S1 shows the contribution (standardized coefficient) of each miRNA in the most optimal prediction model. The performance of the most optimal prediction model on the transformed miRNA expression data was visualized by ROC curve and evaluated by AUC, sensitivity, and specificity. 95% confidence intervals (CI) were calculated for the AUC obtained for the most optimal prediction model[34]. Additionally, we employed the Kruskal-Wallis omnibus test and the *post hoc* non-parametric test to compare miRNA expression between biological groups. Obtained p-values were adjusted for multiple comparisons. All statistical analyses were performed in R open source software using the pROC[35] package.

RESULTS

Discovery of a miRNA Signature in hrHPV-Positive Self-Samples

In order to identify a CIN3-specific miRNA signature in self-samples, we performed genome-wide small RNA sequencing on 74 hrHPV-positive self-samples (discovery set). This yielded an average of 41 million raw reads per sample (controls: 48 million, CIN3: 37 million). After preprocessing (for details please refer to Materials and Methods), we obtained an average of 25 million

reads per sample (controls: 29 million, CIN3: 23 million) of which approximately 24% was mapped to the human genome (controls: 20%, CIN3: 24%). Most non-human reads mapped to Lactobacillus. A fraction of the mapped reads represented miRNAs, with an average of 4% per sample (controls: 5%, CIN3: 3%). Based on quality control assessment (for details please refer to Materials and Methods), 18 samples were excluded resulting in a total of 56 samples (32 controls, 24 CIN3) for further analysis.

To identify a miRNA signature with high discriminatory power for CIN3, we performed adaptive group regularized ridge regression, GRridge[27], which enables objective use of co-data. This resulted in the identification of a panel of nine miRNAs (let-7b, miR-9, miR-15b, miR-20a, miR-31, miR-93, miR-183, miR-184, and miR-222) as strong predictor of CIN3 with an AUC of 0.89 (Fig. 1A). The expression of all miRNAs was found to be increased in CIN3 compared with hrHPV-positive controls (Fig. 1B).

In order to ensure that the identified miRNA markers in self-samples are related to cancer development, we next analyzed in-house available genome-wide miRNA profiles of a set of 37 cervical tissues (10 controls, 18 high-grade CIN, 9 SCC)[30]. In concordance with the expression patterns observed in self-samples,

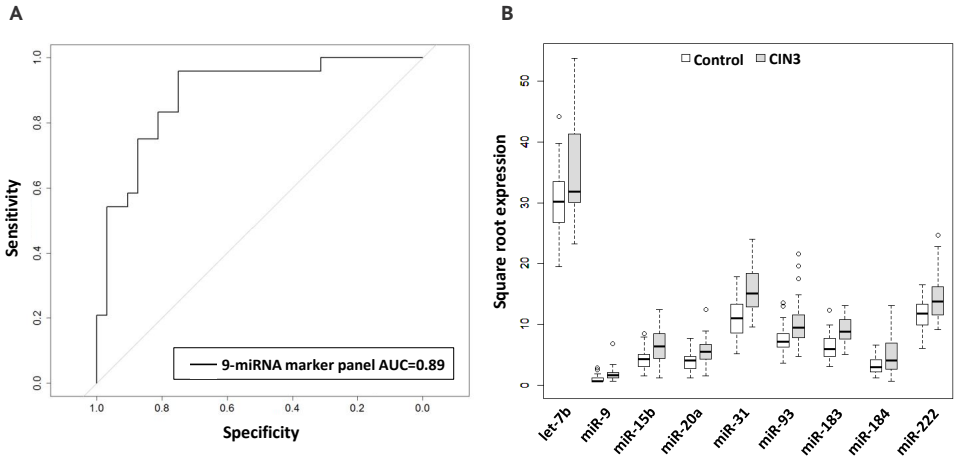

Figure 1. Identification of a CIN3-specific miRNA signature in hrHPV-positive self-samples. (**A**) Receiver operating characteristics (ROC) curve and area under the curve (AUC) of a 9-miRNA marker panel for CIN3 detection in hrHPV-positive self-samples (discovery set; 32 controls and 24 CIN3). (**B**) Square root transformed expression levels of the nine miRNAs in hrHPV-positive self-samples [discovery set; 32 controls (open box) and 24 CIN3 (light grey box)].

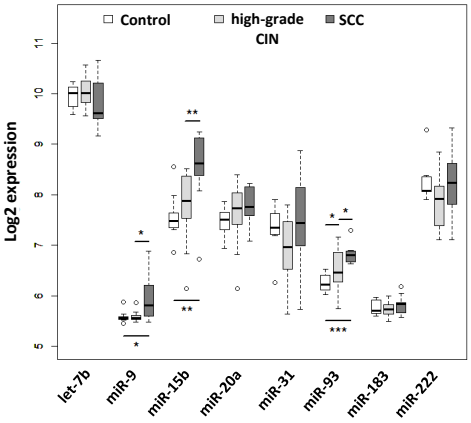

Figure 2. Expression levels of the miRNA signature in cervical tissues. Log2 expression levels obtained from miRNA microarray data of eight out of nine miRNAs in cervical tissues [10 controls (open box), 18 high-grade CIN (light grey box), and 9 SCC (dark grey box)]. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

we observed increased expression for miR-9, miR-15b, miR-20a, miR-93, and miR-183 in cervical tissues from women with high-grade CIN compared with controls (Fig. 2). For let-7b, miR-31, and miR-222, however, we found similar or decreased expression levels in high-grade CIN cervical tissues compared with controls, although these differences were not significant (Fig. 2). Importantly, all miRNAs except let-7b showed increased expression in SCC compared with controls.

Validation of the miRNA Signature by qPCR in hrHPV-Positive Self-Samples

To validate the clinical performance of the nine miRNAs identified in the discovery phase, we performed miRNA qPCR in an independent set of 191 hrHPV-positive self-samples (validation set). Among the nine miRNAs, three miRNAs (miR-9, miR-183, and miR-184) exhibited very low expression levels, detected at an average C_t of > 32. Since the quantification of low expressed transcripts can reduce reproducibility and substantially increase technical PCR noise[31], we excluded these three miRNAs from subsequent analysis.

In concordance with the results presented for the discovery set, we found a significant difference for miR-15b and miR-93 between self-samples from women with CIN3 and hrHPV-positive controls (Fig. 3A). In contrast, for let-7b, we observed significantly lower expression levels in CIN3 self-samples compared with the hrHPV-positive controls, and even lower levels in SCC (Fig. 3A). Interestingly, however, the results for let-7b are in line with the data from cervical tissues for which we also observed decreased expression of let-7b in SCC (Fig. 2). For miR-31 no significant difference between the biological groups

was observed (Fig. 3A). Furthermore, all miRNAs except miR-31 and let-7b showed significantly increased expression levels in self-samples from women with SCC compared with hrHPV-positive controls (Fig. 3A).

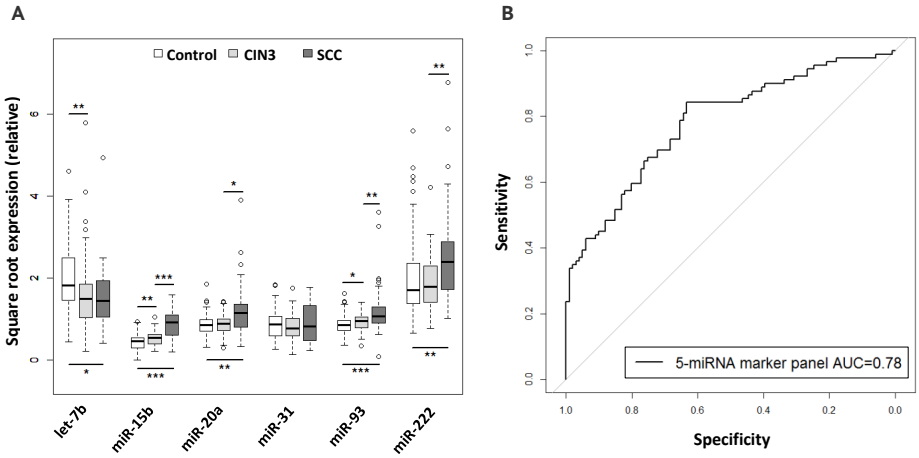

Figure 3. Clinical performance and expression levels of the miRNA markers in hrHPV-positive self-samples. (**A**) Square root transformed expression levels (relative to miR-423 and miR-30b) of the six miRNAs in hrHPV-positive self-samples [validation set; 101 controls (open box), 48 CIN3 (light grey box), and 41 SCC (dark grey box)]. (**B**) ROC curve and AUC of a 5-miRNA marker panel (let-7b, miR-15b, miR-20a, miR-93, and miR-222) in hrHPV-positive self-samples for CIN3+ detection (validation set; 101 controls, 48 CIN3, and 41 SCC). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Next, we performed univariable logistic regression to assess the performance of each miRNA and multivariable logistic regression followed by backward elimination to identify the best combination of miRNA markers. Since we set out to find a panel of miRNA markers that can be applied in cervical screening that should be able to detect both CIN3 and cervical cancer with high discriminatory power, we combined CIN3 and cervical cancer (CIN3+) for the analysis of the validation set. The best clinical performance was achieved by multivariable logistic regression with a panel of five miRNAs (let-7b, miR-15b, miR-20a, miR-93, and miR-222) for which we obtained an AUC of 0.78 (95% confidence interval (CI): 0.7173-0.8479) for CIN3+ detection (Fig. 3B). Expectedly, we observed a better clinical performance for the 5-miRNA marker panel compared with the individual clinical performance of the miRNAs (Fig. S2). We included thresholds corresponding to 65-75% specificity to determine the sensitivity of the 5-miRNA marker panel for CIN3 and SCC

detection (Table 1). At the threshold corresponding to 65% specificity in hrHPV-positive controls, the 5-miRNA marker panel revealed a sensitivity of 67% (32 out of 48) for CIN3 detection. Importantly, 38 out of 41 (93%) of the cervical cancers were detected (Table 1).

Table 1. Clinical performance of the 5-miRNA marker panel in hrHPV-positive self-samples.

Specificity (%)	Sensitivity CIN3 (%)	Sensitivity SCC (%)
75	50	88
70	54	88
65	67	93

DISCUSSION

This is the first study to determine genome-wide miRNA profiles in hrHPV-positive self-samples to allow for direct and objective triage in cervical screening. Since hrHPV-positive self-samples are relatively impure due to the large amount of non-cervical (i.e. vaginal cells, lymphocytes) and normal cervical epithelial cells, we applied our recently proposed GRridge model[27] to identify a miRNA signature with high discriminatory power for CIN3 detection. This method enables objective use of auxiliary data, such as miRNA abundance and conservation, and was previously shown to outperform other prediction methods[29]. We identified a miRNA signature consisting of nine miRNAs (let-7b, miR-9, miR-15b, miR-20a, miR-31, miR-93, miR-183, miR-184, and miR-222) with high discriminatory power for CIN3 detection in hrHPV-positive self-samples (AUC = 0.89).

Importantly, our study shows the feasibility of conducting sRNA sequencing for miRNA quantification on self-samples. It should be noted, however, that prior to the analysis we had to exclude a substantial amount of samples of insufficient quality (18 out of 74). This quality filtering was important to improve the miRNA selection and any potential bias introduced is countered in the validation phase where independent hrHPV-positive self-samples were used without any filtering. Moreover, we found that differential expression of the identified miRNAs could be confirmed in independent cervical tissue specimens analyzed by a different method[30]. In addition, recent data show increased expression of miR-15b in hrHPV-positive physician-taken cervical scrapes from women with CIN3 and SCC[33]. Similarly, preliminary miRNA expression data show consistent

results for miR-20a, miR-93, and miR-222 in hrHPV-positive cervical scrapes (Fig. S3). Taken together, this shows that the detection of the aforementioned miRNAs in self-samples reflects cervical disease and supports the validity of our approach.

Consistent with our findings, Li and colleagues found increased expression of miR-15b, miR-20a, miR-31, miR-93, and miR-222 in cervical precancerous tissues[36]. These miRNAs have also been shown to be upregulated in cervical cancer tissues compared with controls[16,17,37]. Moreover, upregulation of miR-15b was found to be directly linked to a chromosomal gain of 3q, which is often observed in cervical cancers[30]. Additionally, miR-15b expression was recently shown to gradually increase from CIN1 to CIN2/3 and cervical cancer tissues and to be associated with poor prognosis[37].

Validation by qPCR in an independent set of hrHPV-positive self-samples resulted in a good clinical performance for a 5-miRNA signature consisting of let-7b, miR-15b, miR-20a, miR-93, and miR-222 (AUC = 0.78) with a sensitivity of 67% and a specificity of 65% for CIN3 detection. Importantly, 93% (38 out of 41) of the cervical cancers were detected. Since two of the three miRNA-negative cervical cancer self-samples were just below the threshold, this may indicate that these miRNA-negative self-samples do not contain sufficient numbers of representative cells or well-sampled material and warrant additional testing. Compared with previously described molecular triage markers in hrHPV-positive self-samples[9,12,38–40], we obtained a nearly comparable clinical performance for CIN3+ detection as observed for DNA methylation panels[9,38]. Unlike DNA methylation analysis, miRNA quantification by qPCR does not require bisulphite conversion prior to PCR amplification and low amounts of starting material are sufficient for accurate and reproducible analysis. Unfortunately, most miRNA quantification methods only offer assays for each miRNA individually. For future (semi)high-throughput full molecular screening, a combined measurement of several miRNAs into one single test would be desirable. Although we included a multiplex reverse transcription (RT) step that allows for simultaneous transcription of all our miRNAs including reference genes, further technical advancements are warranted, thereby saving self-sampled material, time, and costs.

One limitation of our study that should be considered in interpreting our findings is that we were unable to validate all nine miRNAs by qPCR due to low expression levels of three miRNAs (miR-9, miR-183, and miR-184). This could have affected

the performance of the miRNA signature in the validation set. Also, we applied a different method for validation compared with the discovery (qPCR versus sRNA-Seq). Although several studies have shown close correlations between qPCR and RNA-Seq data[41], differences in miRNA quantification between the methods have been reported[42,43]. A possible explanation for discrepancies between the methods could be the existence of miRNA variants. With the rapid development of sRNA-Seq came the realization that numerous miRNAs exist as multiple length variants, which are termed isomiRs[44]. The majority of length variants are located at the 3'end of the miRNA sequence[44], which could potentially hamper their identification by qPCR[45]. This might explain our discordant results for let-7b between the discovery and validation set. We found significantly increased expression of let-7b in CIN3 compared with controls in the discovery set, while its expression was found downregulated in CIN3 and SCC in the validation set. Interestingly, reduced expression of let-7b in SCC was also observed in cervical tissue specimens, for which we have miRNA microarray data[30]. Since miRNA probes used in microarrays are designed to detect the canonical miRNA sequence as listed in miRBase[46] and do not bind to all isomiRs as detected by sRNA-Seq, these microarray results are likely to better reflect the results obtained with qPCR. The mapping strategy that we applied for the sRNA-Seq data did not allow for the analysis of isomiRs. Therefore, extensive analysis of our sRNA-Seq data is needed in order to elucidate the presence and marker value of isomiRs in self-samples.

Despite the description of various detection methods for efficient and accurate quantification of miRNAs[47], only some of these approaches have been tested for their ability to distinguish between highly similar miRNA sequences. Although these methods possess high sensitivity for specific sequences, they are not adequately specific and closely related isomiRs are also detected[48]. Therefore, further advancements in isomiR detection methods are needed in order to accurately quantify individual isomiRs.

Furthermore, the use of hrHPV-positive self-sample series from a population of non-attending women can be considered as a limitation. Therefore, future confirmation in a regular population-based sample series is warranted. Also, further research on the miRNA signature in hrHPV-positive clinical samples, as well as a comparison to established triage methods including cytology and HPV16/18 genotyping is needed to confirm the value of the miRNA signature for triage.

In conclusion, by genome-wide miRNA profiling on hrHPV-positive self-samples, we identified a CIN3-specific miRNA signature with good performance. Subsequent validation by qPCR on an independent set of hrHPV-positive self-samples showed a good clinical performance of this miRNA signature for CIN3+ detection. Our findings indicate that miRNA analysis on hrHPV-positive self-samples can greatly improve the management of hrHPV-positive women and facilitates the implementation of self-sampling in cervical cancer screening programs.

ACKNOWLEDGEMENTS

The authors thank Paul Eijk for excellent technical assistance.

REFERENCES

1. Arbyn M., Rebolj M., De Kok I.M.C.M., *et al.* 2009. The challenges of organising cervical screening programmes in the 15 old member states of the European Union. *Eur. J. Cancer* 45(15):2671–2678.
2. Rozemeijer K., de Kok I.M.C.M., Naber S.K., *et al.* 2015. Offering Self-Sampling to Non-Attendees of Organized Primary HPV Screening: When Do Harms Outweigh the Benefits? *Cancer Epidemiol. Biomarkers Prev.* 24(5):773–782.
3. Gök M., Heideman D.A.M., van Kemenade F.J., *et al.* 2010. HPV testing on self collected cervicovaginal lavage specimens as screening method for women who do not attend cervical screening: cohort study. *BMJ* 340:c1040.
4. Arbyn M., Verdoodt F., Snijders P.J.F., *et al.* 2014. Accuracy of human papillomavirus testing on self-collected versus clinician-collected samples: a meta-analysis. *Lancet Oncol.* 15(2):172–183.
5. Arbyn M. & Castle P.E. 2015. Offering self-sampling kits for HPV testing to reach women who do not attend in the regular cervical cancer screening program. *Cancer Epidemiol. Biomarkers Prev.* 24(5):769–772.
6. Rijkaart D.C., Berkhof J., van Kemenade F.J., *et al.* 2012. Evaluation of 14 triage strategies for HPV DNA-positive women in population-based cervical screening. *Int. J. Cancer* 130(3):602–610.
7. Dijkstra M.G., van Niekerk D., Rijkaart D.C., *et al.* 2014. Primary hrHPV DNA Testing in Cervical Cancer Screening: How to Manage Screen-Positive Women? A POBASCAM Trial Substudy. *Cancer Epidemiol. Biomarkers Prev.* 23(1):55–63.
8. Yoshida T., Sano T., Takada N., *et al.* 2011. Comparison of Self-Collected and Clinician-Collected Materials for Cervical Cytology and Human Papillomavirus Genotyping: Analysis by Linear Array Assay. *Acta Cytol.* 55(1):106–112.
9. Verhoef V.M.J., Bosgraaf R.P., van Kemenade F.J., *et al.* 2014. Triage by methylation-marker testing versus cytology in women who test HPV-positive on self-collected cervicovaginal specimens (PROHTECT-3): a randomised controlled non-inferiority trial. *Lancet. Oncol.* 15(3):315–322.
10. Verhoef V.M.J., Heideman D.A.., van Kemenade F.J., *et al.* 2014. Methylation marker analysis and HPV16/18 genotyping in high-risk HPV positive self-sampled specimens to identify women with high grade CIN or cervical cancer. *Gynecol. Oncol.* 135(1):58–63.
11. Ebisch R.M.F., de Kuyper-de Ridder G.M., Bosgraaf R.P., *et al.* 2016. The clinical value of HPV genotyping in triage of women with high-risk-HPV-positive self-samples. *Int. J. Cancer* 139(3):691–699.
12. Hesselink A.T., Heideman D.A.., Steenbergen R.D.., *et al.* 2014. Methylation marker analysis of self-sampled cervico-vaginal lavage specimens to triage high-risk HPV-positive women for colposcopy. *Int. J. Cancer* 135:880–886.
13. Verlaat W., Snoek B.C., Heideman D.A.M., *et al.* 2018. Identification and Validation of a 3-Gene Methylation Classifier for HPV-Based Cervical Screening on Self-Samples. *Clin. Cancer Res.* 24(14):3456–3464.
14. MacFarlane L.-A. & Murphy P.R. 2010. MicroRNA: Biogenesis, Function and Role in Cancer. *Curr. Genomics* 11(7):537–561.
15. Tian Q., Li Y., Wang F., *et al.* 2014. MicroRNA Detection in Cervical Exfoliated Cells as a Triage for Human Papillomavirus – Positive Women. *J. Natl. Cancer Inst.* 106(9):dju241.
16. Satapathy S., Batra J., Jeet V., *et al.* 2017. MicroRNAs in HPV associated cancers: small players with big consequences. *Expert Rev. Mol. Diagn.* 17(7):711–722.
17. Wilting S.M., Verlaat W., Jaspers A., *et al.* 2013. Methylation-mediated transcriptional repression of microRNAs during cervical carcinogenesis. *Epigenetics* 8(2):220–228.
18. Mo W., Tong C., Zhang Y., *et al.* 2015. microRNAs' differential regulations mediate the progress of Human Papillomavirus (HPV)-induced Cervical Intraepithelial Neoplasia (CIN). *BMC Syst. Biol.* 9:4.

19. Herrington C.S. 2015. The terminology of pre-invasive cervical lesions in the UK cervical screening programme. *Cytopathology* 26(6):346–350.
20. Bosgraaf R.P., Verhoef V.M.J., Massuger L.F.A.G., *et al.* 2015. Comparative performance of novel self-sampling methods in detecting high-risk human papillomavirus in 30,130 women not attending cervical screening. *Int. J. Cancer* 136(3):646–655.
21. Hesselink A.T., Berkhof J., van der Salm M.L., *et al.* 2014. Clinical validation of the HPV-risk assay, a novel real-time PCR assay for detection of high-risk human papillomavirus DNA by targeting the E7 region. *J. Clin. Microbiol.* 52(3):890–896.
22. Federa. (2011) Federa. Human Tissue and Medical Research: Code of conduct for responsible use (2011). Available at: http://www.federa.org/sites/default/files/digital_version_first_part_code_of_conduct_in_uk_2011_12092012.pdf [Accessed on 15 October 2106]
23. Friedländer M.R., Chen W., Adamidi C., *et al.* 2008. Discovering microRNAs from deep sequencing data using miRDeep. *Nat. Biotechnol.* 26(4):407–415.
24. Friedländer M.R., Mackowiak S.D., Li N., *et al.* 2012. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 40(1):37–52.
25. Kozomara A. & Griffiths-Jones S. 2011. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* 39(Database issue):D152–D157.
26. Robinson M.D.M., Oshlack A., Wang E., *et al.* 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11(3):R25.
27. van de Wiel M.A., Lien T.G., Verlaat W., *et al.* 2016. Better prediction by use of co-data: Adaptive group-regularized ridge regression. *Stat. Med.* 35(3):368–381.
28. Agarwal V., Bell G.W., Nam J.-W., *et al.* 2015. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 4.
29. Novianti P.W., Snoek B.C., Wilting S.M., *et al.* 2017. Better diagnostic signatures from RNAseq data through use of auxiliary co-data. *Bioinformatics* 33:1572–1574.
30. Wilting S.M., Snijders P.J.F., Verlaat W., *et al.* 2013. Altered microRNA expression associated with chromosomal changes contributes to cervical carcinogenesis. *Oncogene* 32(1):106–116.
31. Babion I., Snoek B.C., van de Wiel M.A., *et al.* 2017. A Strategy to Find Suitable Reference Genes for miRNA Quantitative PCR Analysis and Its Application to Cervical Specimens. *J. Mol. Diagnostics* 19(5):625–637.
32. Livak K.J. & Schmittgen T.D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25:402–408.
33. Babion I., Snoek B.C., Novianti P.W., *et al.* 2018. Triage of high-risk HPV-positive women in population-based screening by miRNA expression analysis in cervical scrapes; a feasibility study. *Clin. Epigenetics* 10:76.
34. DeLong E.R., DeLong D.M. & Clarke-Pearson D.L. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44(3):837–845.
35. Robin X., Turck N., Hainard A., *et al.* 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12(1):77.
36. Li Y., Wang F., Xu J., *et al.* 2011. Progressive miRNA expression profiles in cervical carcinogenesis and identification of HPV-related target genes for miR-29*. *J. Pathol.* 224(4):484–495.
37. Wen F., Xu J.-Z. & Wang X.-R. 2017. Increased expression of miR-15b is associated with clinicopathological features and poor prognosis in cervical carcinoma. *Arch. Gynecol. Obstet.* 295(3):743–749.
38. De Strooper L.M.A., Verhoef V.M.J., Berkhof J., *et al.* 2016. Validation of the FAM19A4 / mir124-2 DNA methylation test for both lavage- and brush-based self-samples to detect cervical (pre) cancer in HPV-positive women. *Gynecol. Oncol.* 141(2):341–347.

39. Luttmer R., De Strooper L.M.A., Steenbergen R.D.M., *et al.* 2016. Management of high-risk HPV-positive women for detection of cervical (pre)cancer. *Expert Rev. Mol. Diagn.* 16(9):961–974.
40. Steenbergen R.D.M., Snijders P.J.F., Heideman D.A.M., *et al.* 2014. Clinical implications of (epi) genetic changes in HPV-induced cervical precancerous lesions. *Nat. Rev. Cancer* 14(6):395–405.
41. Git A., Dvinge H., Salmon-Divon M., *et al.* 2010. Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. *RNA* 16(5):991–1006.
42. Koppers-Lalic D., Hackenberg M., de Menezes R., *et al.* 2016. Non-invasive prostate cancer detection by measuring miRNA variants (isomiRs) in urine extracellular vesicles. *Oncotarget* 7(16):22566–22578.
43. Sapre N., Hong M.K.H., Macintyre G., *et al.* 2014. Curated MicroRNAs in Urine and Blood Fail to Validate as Predictive Biomarkers for High-Risk Prostate Cancer. *PLoS One* 9(4):e91729.
44. Neilsen C.T., Goodall G.J. & Bracken C.P. 2012. IsomiRs – the overlooked repertoire in the dynamic microRNAome. *Trends Genet.* 28(11):544–549.
45. Schamberger A. & Orbán T.I. 2014. 3′ IsomiR Species and DNA Contamination Influence Reliable Quantification of MicroRNAs by Stem-Loop Quantitative PCR. *PLoS One* 9(8):e106315.
46. Kozomara A. & Griffiths-Jones S. 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42(D1):D68–D73.
47. Mestdagh P., Hartmann N., Baeriswyl L., *et al.* 2014. Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study. *Nat. Methods* 11(8):809–815.
48. Magee R., Telonis A.G., Cherlin T., *et al.* 2017. Assessment of isomiR Discrimination Using Commercial qPCR Methods. *Non-coding RNA* 3(2):18.

SUPPLEMENTARY MATERIALS

Table S1. Contribution of each miRNA indicated by the standardized coefficient in the most optimal prediction model.

miRNA	Standardized coefficient
miR-222	-0.350
miR-20a	-0.952
let-7b	-1.482
miR-93	1.731
miR-15b	3.767

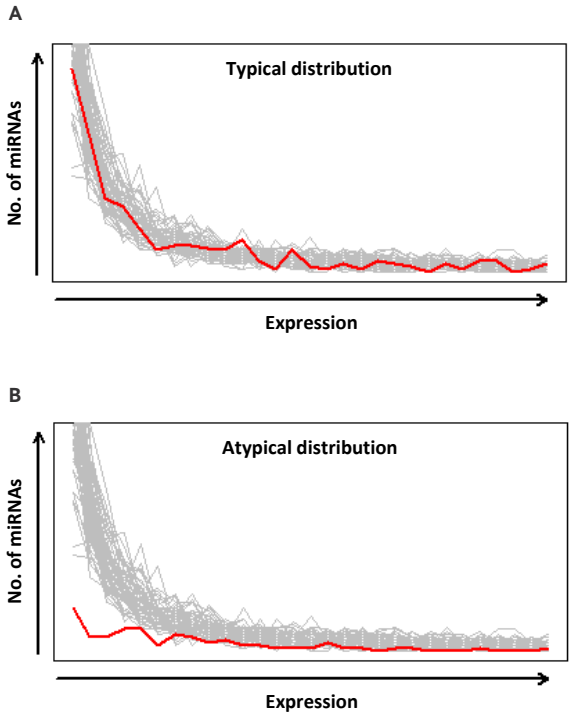

Figure S1. Distribution of total miRNA read counts in hrHPV-positive self-samples. Self-samples with a typical miRNA read count distribution are depicted in grey. As example, (**A**) one self-sample with a typical miRNA read count distribution and (**B**) one self-sample with an atypical miRNA read count distribution are depicted in red.

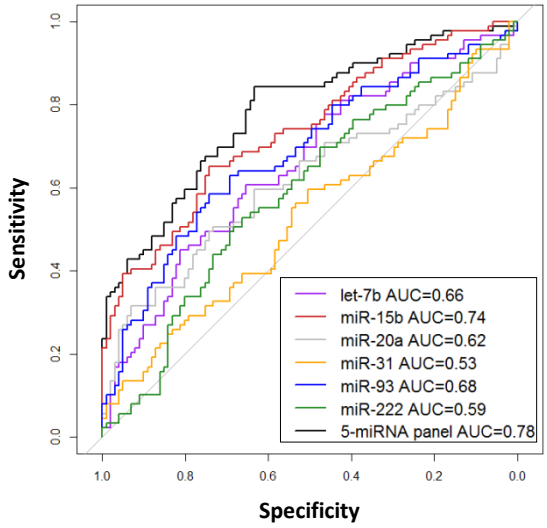

Figure S2. Individual clinical performance of the six miRNAs in hrHPV-positive self-samples. ROC curve and AUC of the six miRNAs in hrHPV-positive self-samples (validation set; 101 controls, 48 CIN3, and 41 SCC). For comparison, the ROC curve and AUC of the 5-miRNA marker panel is included.

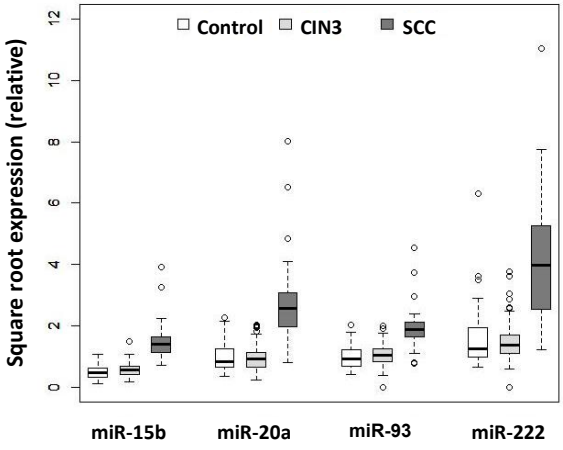

Figure S3. Expression levels of miRNAs in hrHPV-positive cervical scrapes. Square root transformed expression levels (relative to RNU24 and miR-423) of miR-15b, miR-20a, miR-93, and miR-222 in hrHPV-positive cervical scrapes from 60 women without cervical disease (white box), 119 women with CIN3 (light grey box), and 29 women with SCC (dark grey box).